

# Bioinformatics

Week 1, 2020-03-04

Xiaofan Zhou

#		#	#####	#		#####	#####	#		#	#####	###
#	#	#	#	#		#	#	#	##	##	#	###
#	#	#	#	#		#	#	#	#	#	#	###
#	#	#	#####	#		#	#	#	#	#	#####	#
#	#	#	#	#		#	#	#	#	#	#	
#	#	#	#	#		#	#	#	#	#	#	###
	##	##	#####	#####		#####	#####	#		#	#####	###

# **Week 1: Introduction to ...**

- The participants
- The course
- The discipline

# Who am I?

- Xiaofan Zhou (周筱帆)
  - Professor, Integrative Microbiology Research Centre, SCAU
- Research interests:
  - molecular evolution, genomics, and other things
- Education:
  - 2001 – 2005: B.S., Biotechnology, SJTU
  - 2006 – 2011: Ph.D., Cell and Developmental Biology, PSU
- Experience:
  - 2011 – 2016: Postdoc, Vanderbilt University
  - 2016 – Present: Professor, SCAU

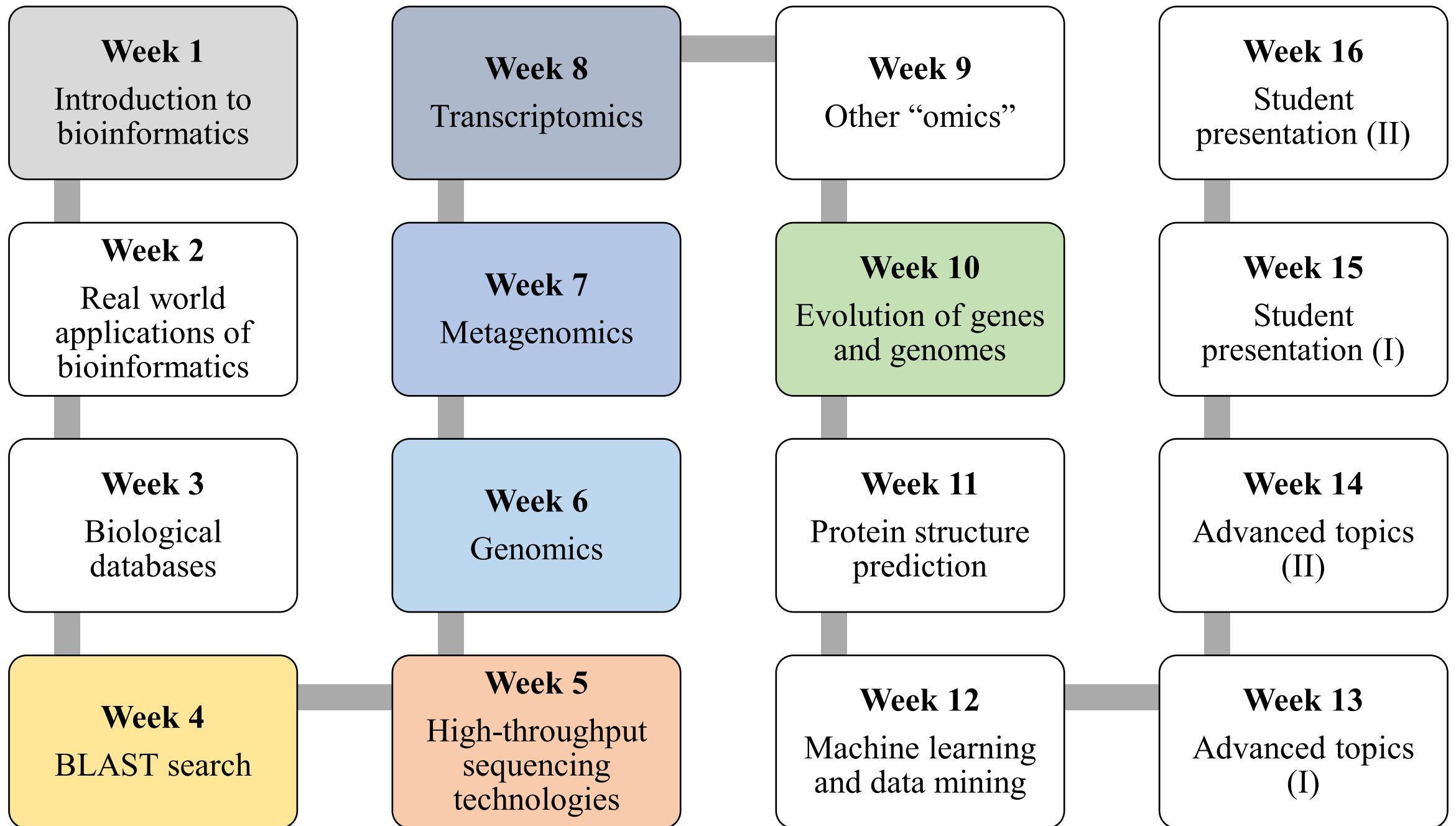
# Week 1: Introduction to ...

- The participants
- The course



# Course overview

- Prerequisite:
  - basic biology knowledge and computer skills
  - personal computer (for the practical session)
- Aim:
  - to introduce core concepts and methods in bioinformatics
- Organization:
  - coordinated lecture and practical sessions



# Course overview

- Prerequisite:
  - basic biology knowledge and computer skills
  - personal computer (for the practical session)
- Aim:
  - to introduce core concepts and methods in bioinformatics
- Organization:
  - alternate lecture and practical session
- Grading policy:
  - 50%: attendance, homework, presentation
  - 50%: final exam



# Week 1: Introduction to ...

- The participants
- The course
- The discipline

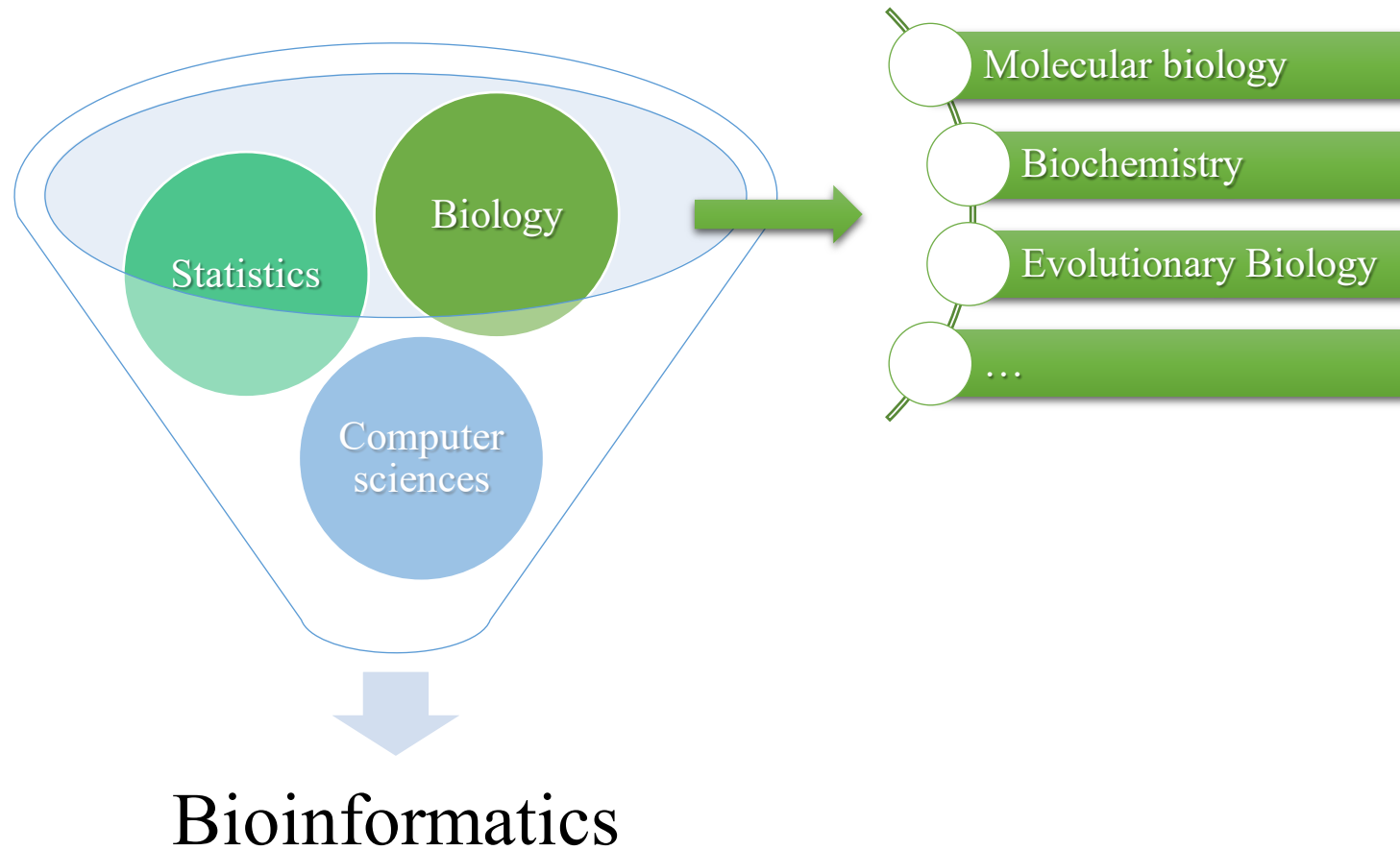
# What is bioinformatics?

- NIH working definition:
  - Bioinformatics: *“Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.”*

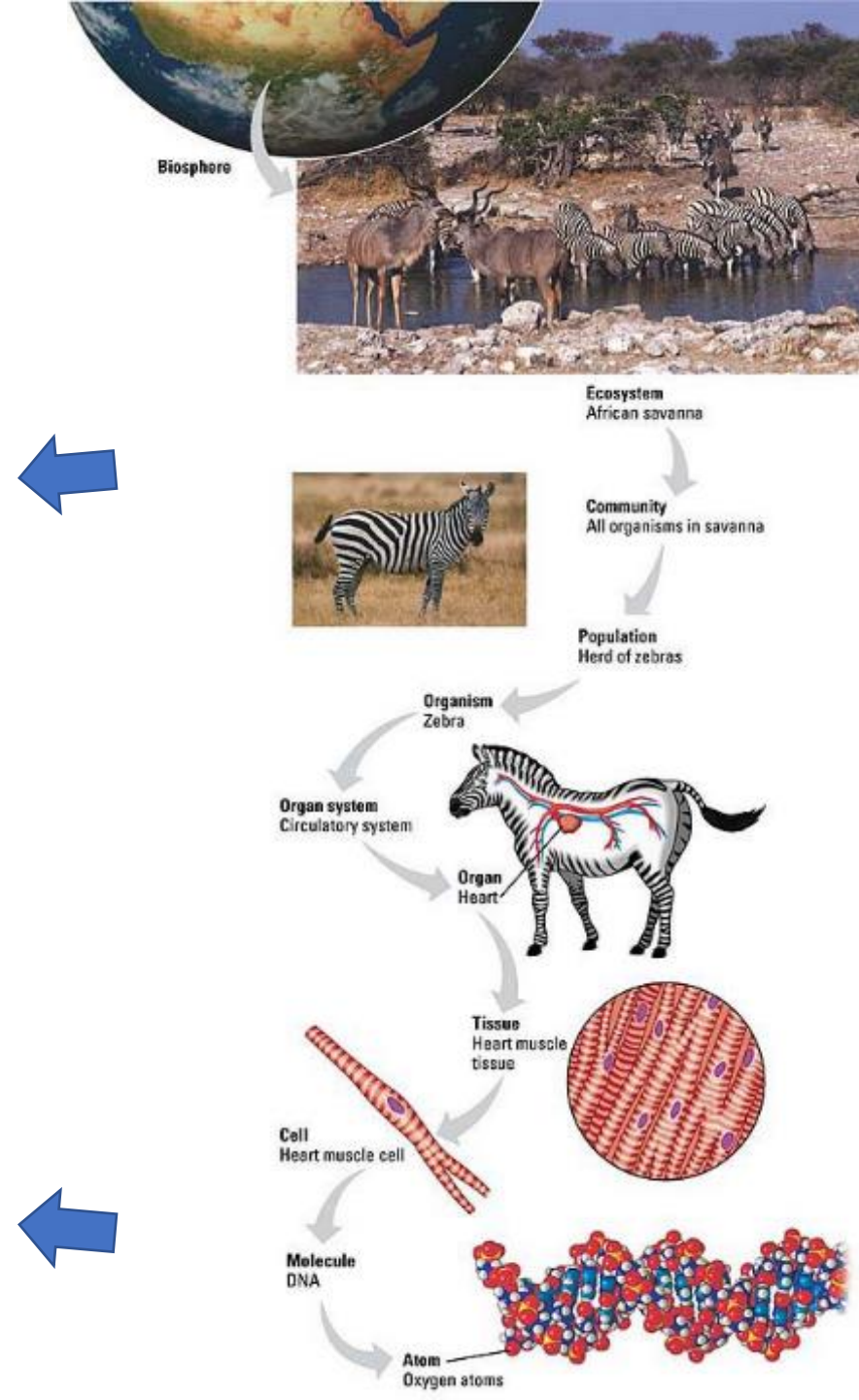
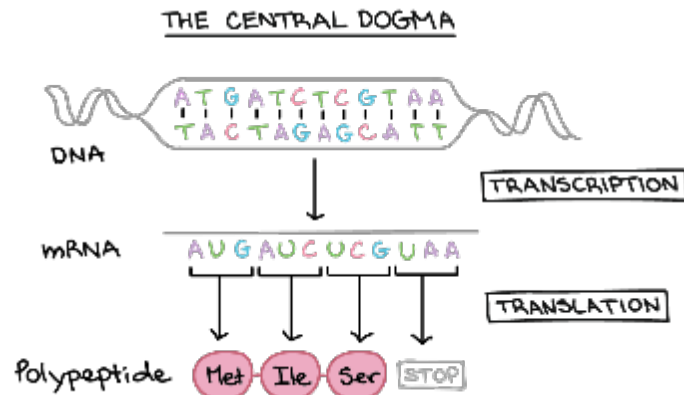


**“Application of computer to biology”**

# Core components of bioinformatics

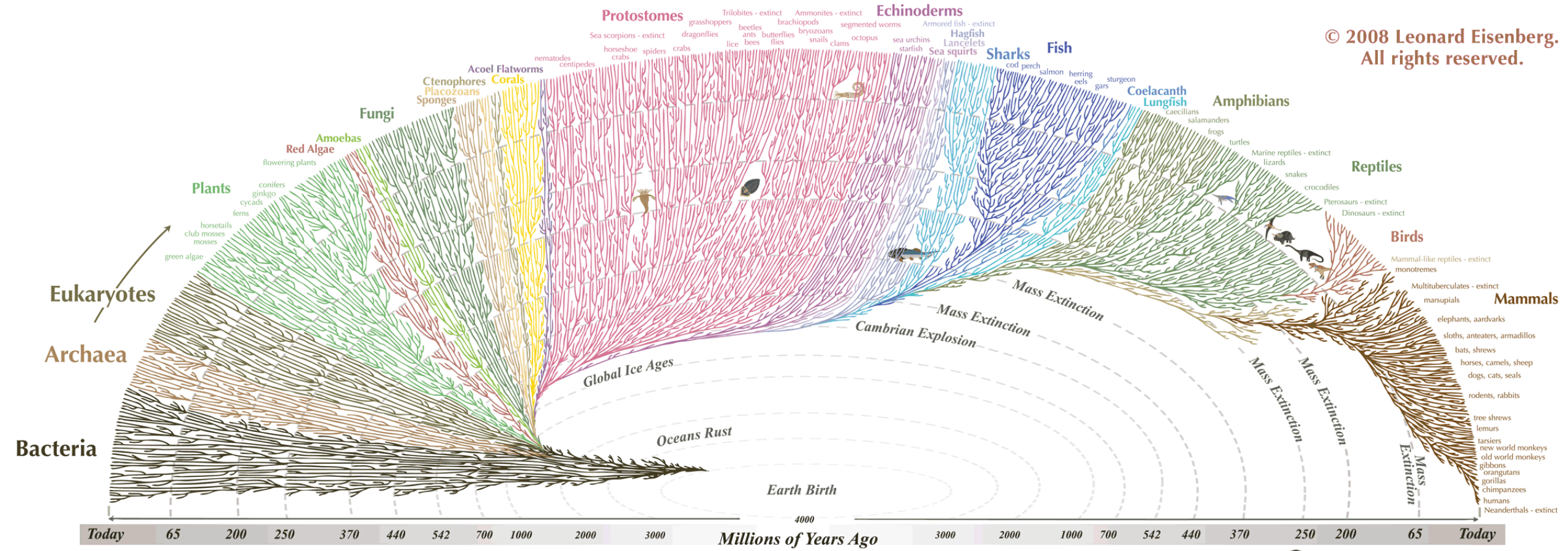


# From micro to macro



# From past to present

© 2008 Leonard Eisenberg.  
All rights reserved.



All the major and many of the minor living branches of life are shown on this diagram, but only a few of those that have gone extinct are shown. Example: Dinosaurs - extinct



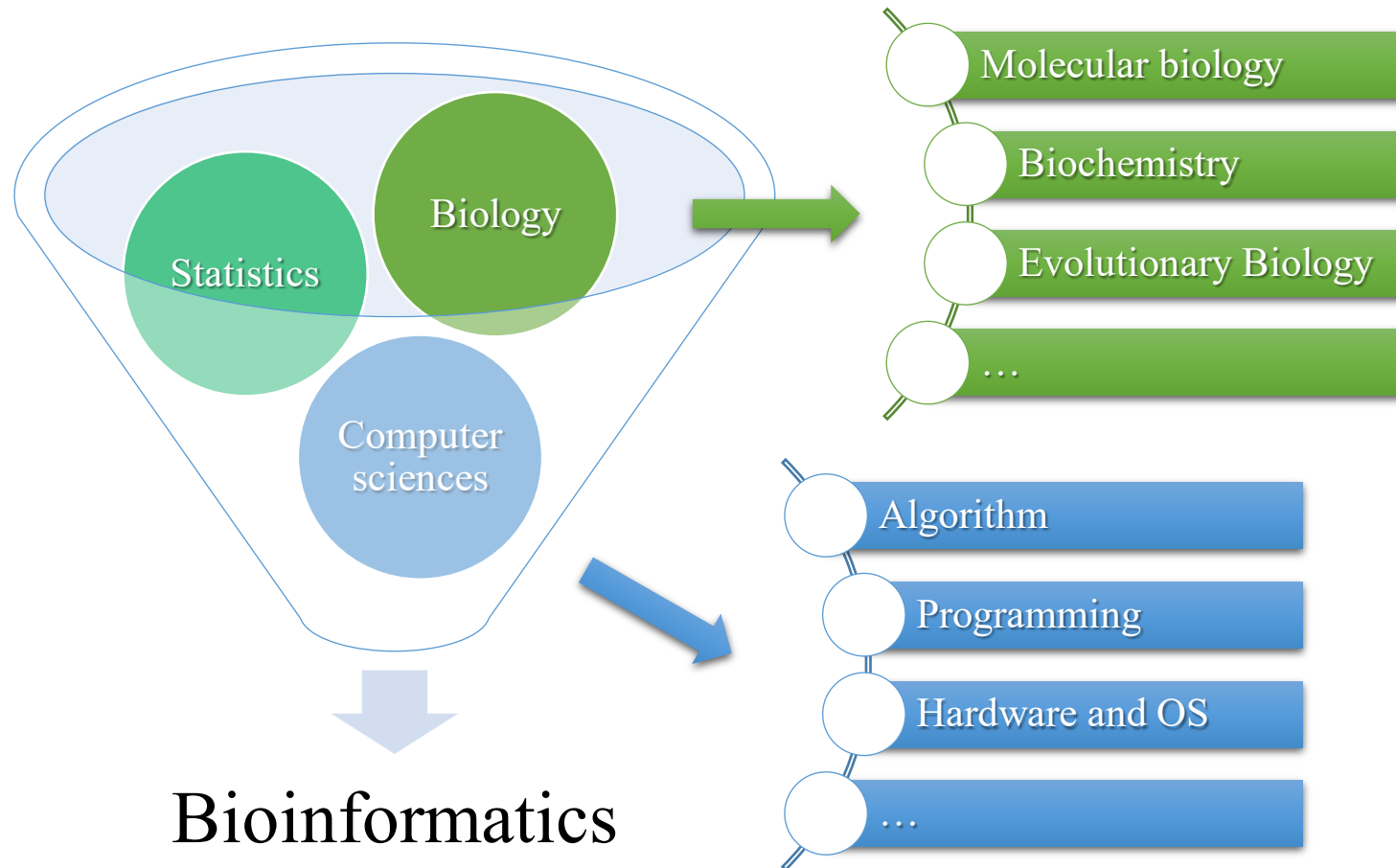
© 2008 Leonard Eisenberg. All rights reserved.  
evogeneao.com



# Fighting the coronavirus



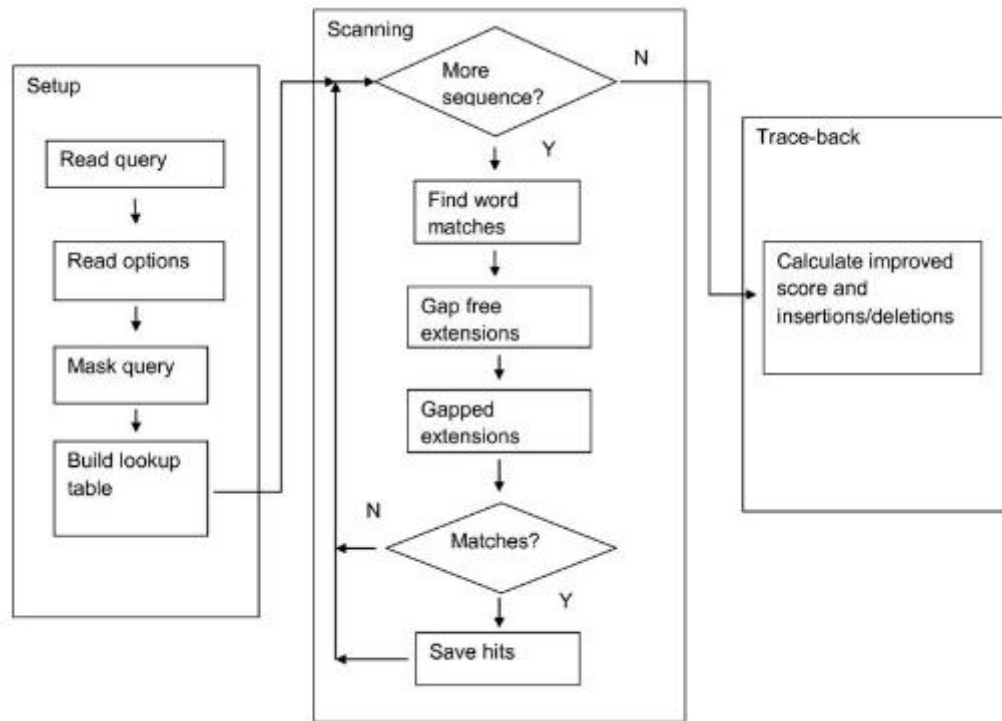
# Core components of bioinformatics



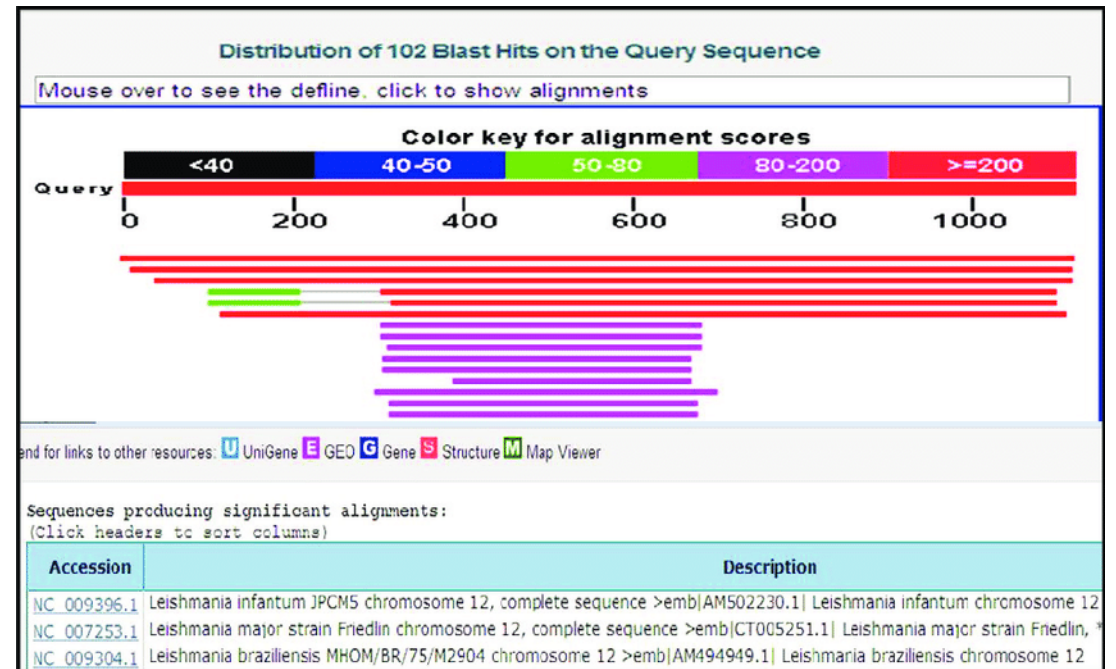


# “Hardcore” vs. “softcore” bioinformatics

## Making tools



## Using tools



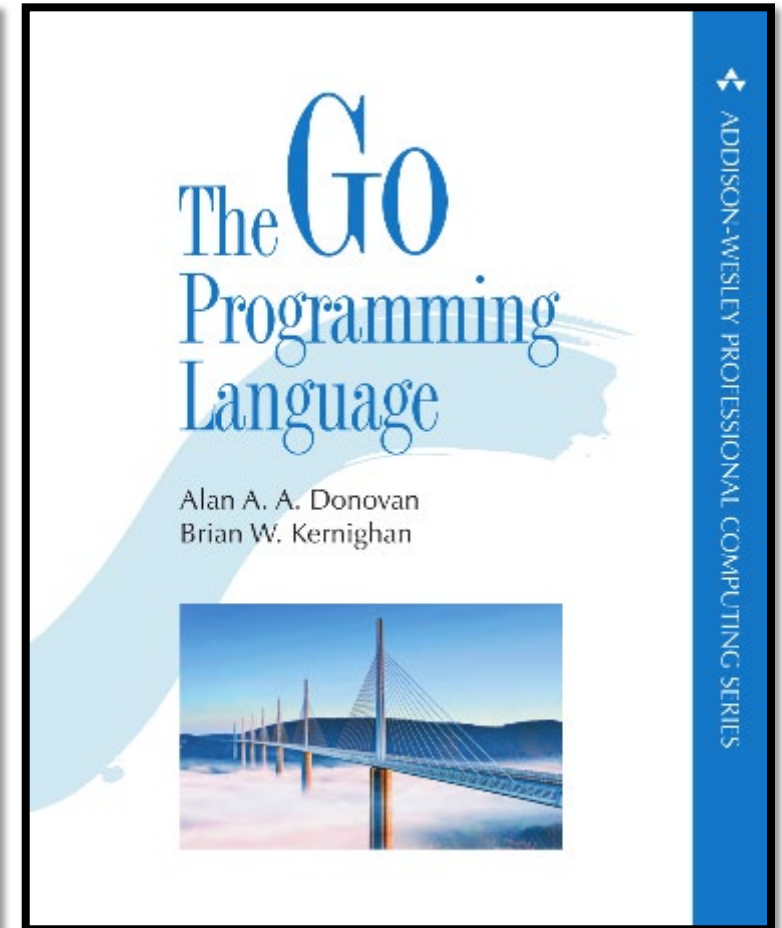
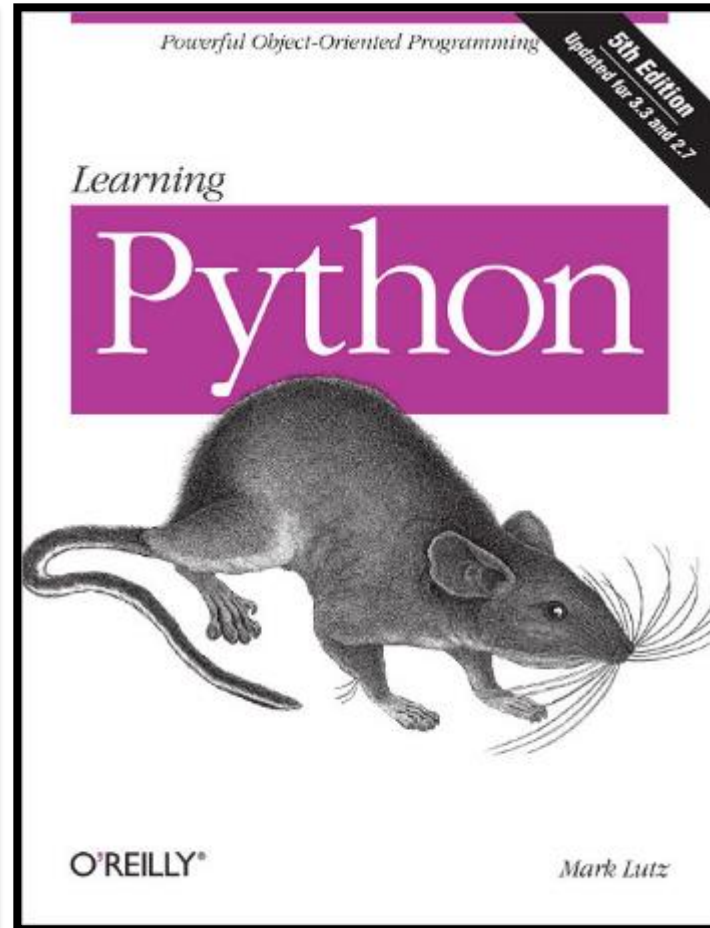
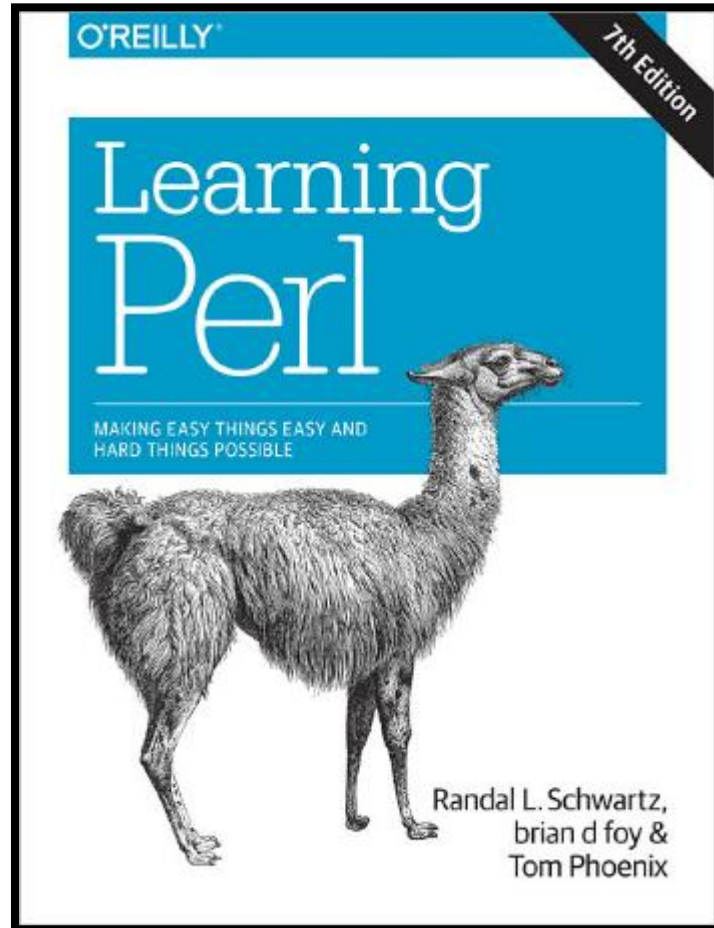
# Bioinformatics vs. Computation Biology

- NIH working definition:
  - Bioinformatics: *“Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.”*
  - Computational Biology: *“The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.”*

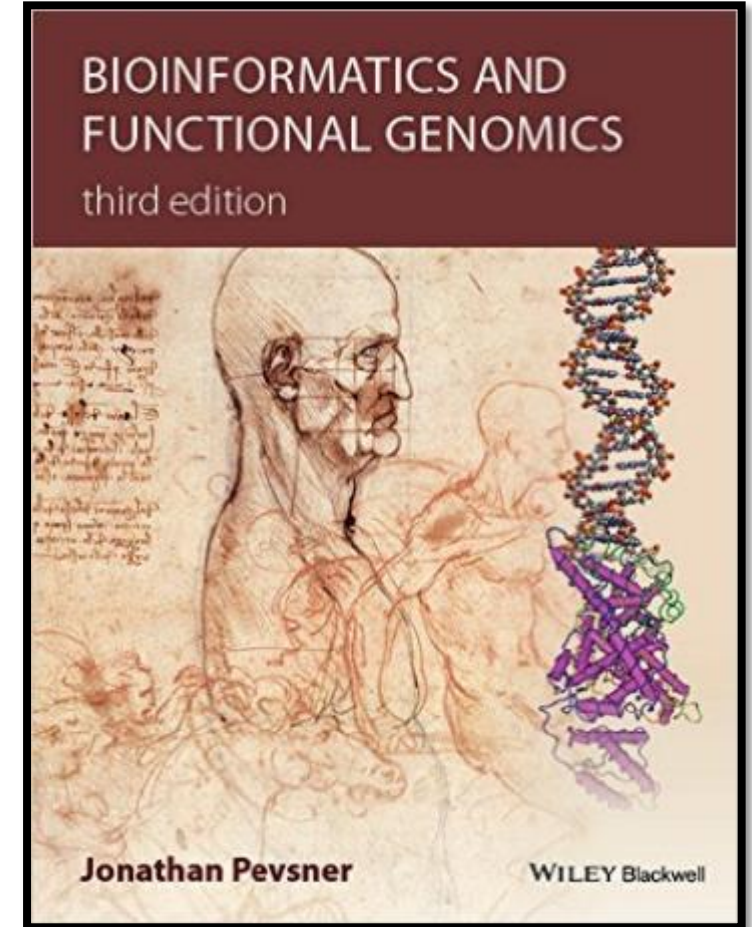
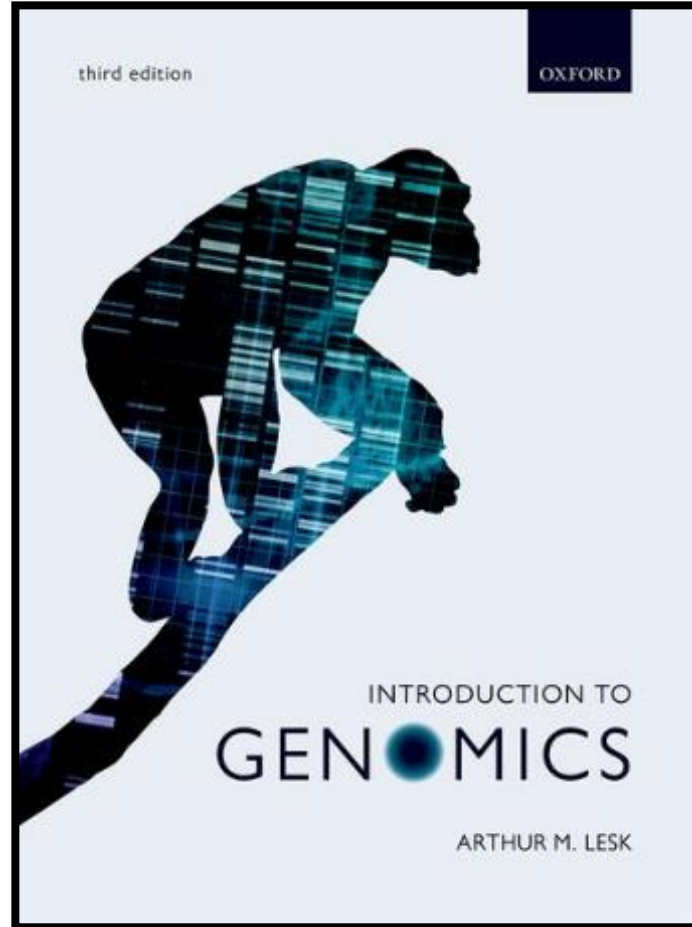
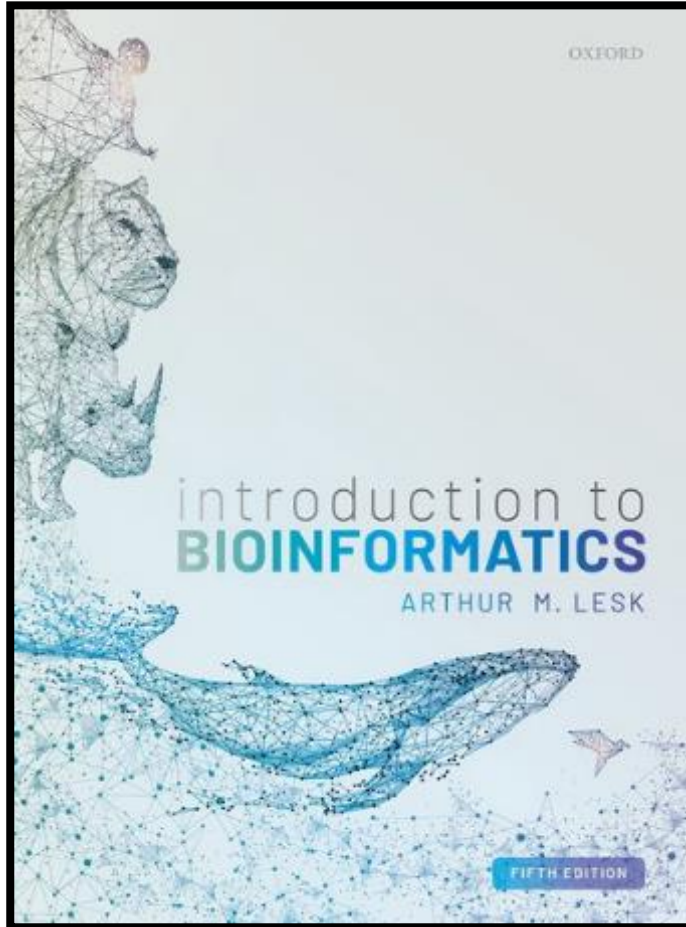
# Self-learning resources

- Books
- Open courses
- Online platforms
- Journals

# Programming language



# Bioinformatics & Genomics





# Open courses

The screenshot displays the Coursera interface for a specific course. At the top, the Coursera logo is on the left, followed by a navigation menu with '探索' (Explore) and a search bar containing '您想学习什么?' (What do you want to learn?). On the right of the header are links for '企业版' (Enterprise), '登录' (Log in), and a blue '免费加入' (Join for free) button.

The main content area has a dark red background. It features a breadcrumb trail: '浏览 > 健康 > 医学信息学' (Browse > Health > Medical Informatics). The course title '生物信息学: 导论与方法' (Bioinformatics: Introduction and Methods) is prominently displayed in white. Below the title is a star rating of 4.7 from 157 reviews and 50 audits. To the right, the provider '提供方' (Provider) is listed as 'PEKING UNIVERSITY' with its logo.

A white box on the left contains the text '免费注册' (Free registration) and '于 Sep 04 开始' (Starting Sep 04). Next to it, the text '有助学金' (Financial aid available) is shown. Below these, it states '13,964 人已注册!' (13,964 people have registered!).

The footer is a light gray bar with links for '关于' (About), '授课大纲' (Syllabus), '审阅' (Reviews), '讲师' (Instructors), '注册选项' (Registration options), and '常见问题解答' (FAQ).

<https://zh.coursera.org/learn/sheng-wu-xin-xi-xue>

# Open courses

The screenshot displays the Coursera website interface for the Bioinformatics Specialization. At the top, the Coursera logo is on the left, followed by an 'Explore' button with a dropdown arrow. A search bar contains the text 'What do you want to learn?' with a magnifying glass icon. To the right of the search bar are links for 'For Enterprise', 'Log In', and a 'Join for Free' button.

The main content area has a dark blue header with the breadcrumb 'Browse > Health > Health Informatics'. Below this, the title 'Bioinformatics Specialization' is prominently displayed. A descriptive text reads: 'Journey to the Frontier of Computational Biology. Master bioinformatics software and computational approaches in modern biology.' To the left of this text is a white button that says 'Enroll for Free' with 'Starts Sep 04' underneath. To the right of the button, it says 'Financial aid available'.

Below the enrollment information, it states '12,546 already enrolled!'. On the right side of the main content area, under the heading 'Offered By', is the text 'UC San Diego'. Below this, a list of six course topics is shown: I. Finding Hidden Messages in DNA, II. Genome Sequencing, III. Comparing Genes, Proteins, and Genomes, IV. Molecular Evolution, V. Genomic Data Science and Clustering, and VI. Finding Mutations in DNA and Proteins.

The footer contains a row of links: 'About', 'How It Works', 'Courses', 'Instructors', 'Enrollment Options', and 'FAQ'.

<https://www.coursera.org/specializations/bioinformatics>



# Open courses



<http://evomics.org/>

# Open courses

EVOLUTION AND GENOMICS					WORKSHOPS	LEARNING	PEOPLE	APPLY	INFORMATION
Intensive and immersive training opportunities									
DATE	DAY	TIME	PRESENTER	TOPIC	LOCATION				
6 Jan	Sunday	6p – 10p	<a href="#">Everyone</a>	Reception	Hotel Zlaty Andel				
7 Jan	Monday	9a – 12p	<a href="#">Scott Handley</a>	<a href="#">Introduction and Orientation</a>	Theater				
	Monday	2p – 5p	<a href="#">Sonya Dyhrman</a>	From flask to field: tracking the drivers of phytoplankton physiological ecology across marine ecosystems <a href="#">Part A</a> and <a href="#">Part B</a>	Theater				
	Monday	7:30p – 10p	Everyone	Scientific Speed Networking	Minorite Monastery				
8 Jan	Tuesday	9a – 12p	<a href="#">Mike Zody</a>	<a href="#">Sequencing Technology &amp; Study Design</a>	Theater				
	Tuesday	2p – 5p	<a href="#">Sophie Shaw</a>	<a href="#">Introduction to Unix and Software Installation Lab</a>	House of Prelate				
	Tuesday	7p – 10p	<a href="#">Hannah Tavalire</a>	<a href="#">Introduction to R</a>	House of Prelate				
9 Jan	Wednesday	9a – 12p	<a href="#">Sophie Shaw</a>	<a href="#">Working with Sequence Data Lab: Quality Control</a>	House of Prelate				
	Wednesday	2p – 5p	<a href="#">Mike Zody</a>	<ul style="list-style-type: none"><li><a href="#">Intro Read Alignment 2019</a></li><li><a href="#">Alignment Workshop</a></li></ul>	House of Prelate				
	Wednesday	7p – 10p	<a href="#">Josephine Paris</a>	<a href="#">Genomics Lab</a>	House of Prelate				




# Online platform

The screenshot shows the 'Locations' page of the Rosalind platform. The header includes the 'ROSALIND' logo, navigation links for 'About', 'Problems', 'Statistics', and 'Glossary', a search bar, and social media icons for Facebook and Twitter. The main content area is titled 'Locations' and contains a paragraph about the platform. Below this, there are five featured sections, each with a red icon, a title, and a description:

- Python Village**: Represented by the Python logo. Description: 'If you are completely new to programming, try these initial problems to learn a few basics about the Python programming language. You'll get familiar with the operations needed to start solving bioinformatics challenges in the Stronghold.'
- Bioinformatics Stronghold**: Represented by a red cloud-like icon. Description: 'Discover the algorithms underlying a variety of bioinformatics topics: computational mass spectrometry, alignment, dynamic programming, genome assembly, genome rearrangements, phylogeny, probability, string algorithms and others.'
- Bioinformatics Armory**: Represented by a red monitor icon with a pair of scissors. Description: 'Ready-to-use software tools abound for bioinformatics analysis. Whereas in the Stronghold you implement algorithms on your own, in the Armory you solve similar problems by using existing tools.'
- Bioinformatics Textbook Track**: Represented by a red book icon. Description: 'A collection of exercises to accompany Bioinformatics Algorithms: An Active-Learning Approach by Phillip Compeau & ...'
- Algorithmic Heights**: Represented by a red tree diagram icon. Description: 'A collection of exercises in introductory algorithms to accompany "Algorithms", the popular textbook by Dasgupta.'

<http://rosalind.info/problems/locations/>

# Online platform

About ▾ Problems ▾ Statistics ▾ Glossary    Log in Register

## Problems

Bioinformatics Stronghold ▾ List Tree

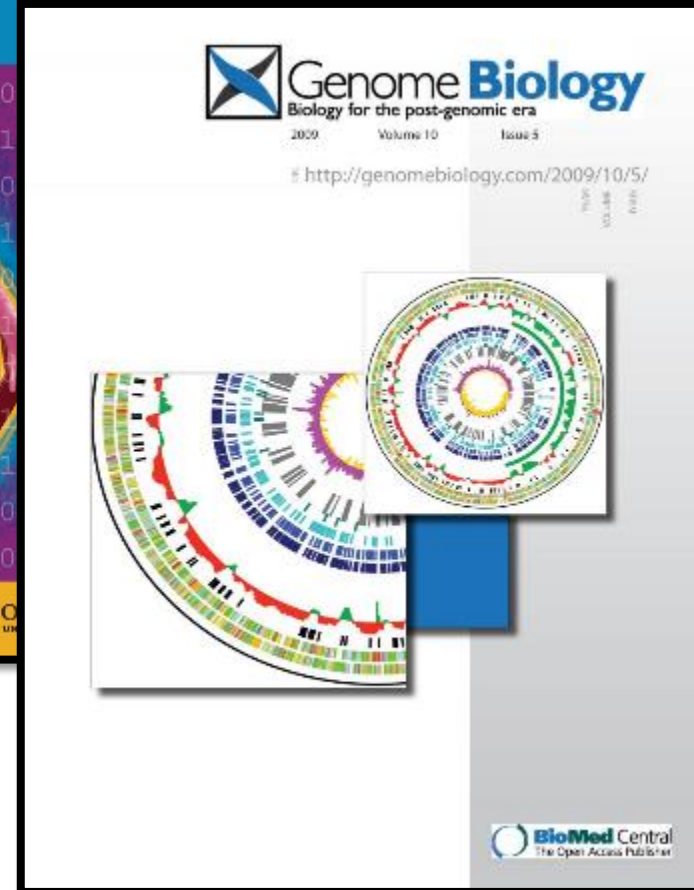
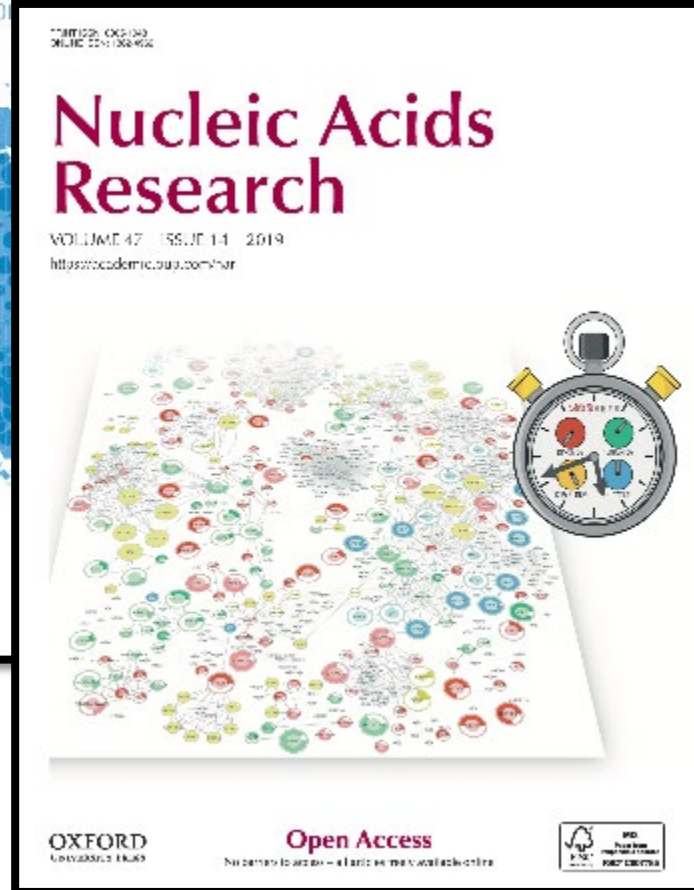
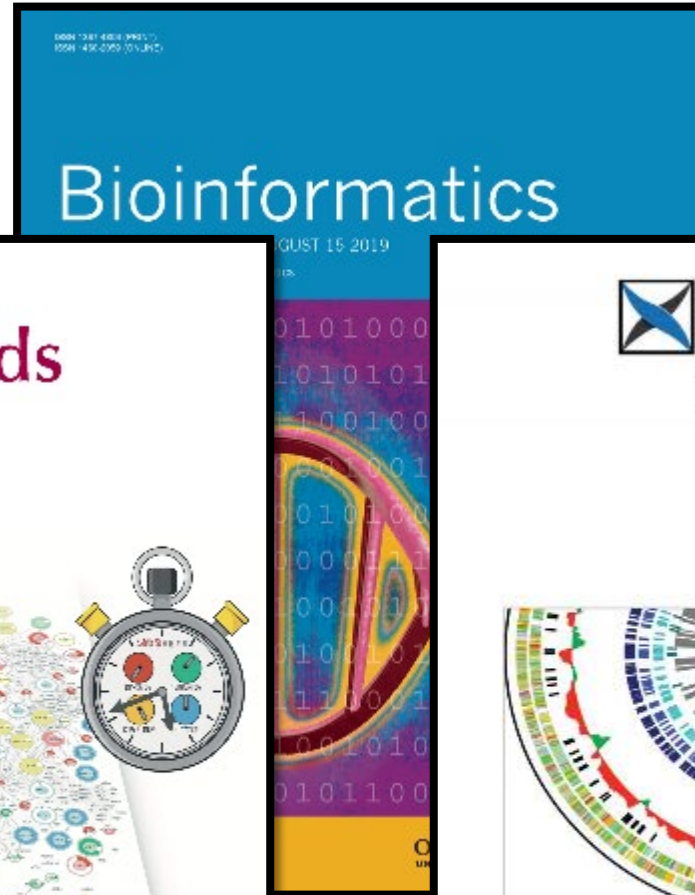
Rosalind is a platform for learning bioinformatics and programming through problem solving. [Take a tour](#) to get the hang of how Rosalind works.

Last win: [victor.mouquin](#) vs. "[Consensus and Profile](#)", 8 minutes ago Problems: 285 (total), users: 66774, attempts: 1104204, correct: 615160

ID	Title	Solved By	Correct Ratio
DNA	Counting DNA Nucleotides	38495	<div><div></div></div>
RNA	Transcribing DNA into RNA	34383	<div><div></div></div>
REVC	Complementing a Strand of DNA	31149	<div><div></div></div>
FIB	Rabbits and Recurrence Relations	17856	<div><div></div></div>
GC	Computing GC Content	18206	<div><div></div></div>
HAMM	Counting Point Mutations	20600	<div><div></div></div>
IPRB	Mendel's First Law	11896	<div><div></div></div>
PROT	Translating RNA into Protein	15993	<div><div></div></div>
SUBS	Finding a Motif in DNA	16414	<div><div></div></div>



# Journals



# Finally...



## 无法访问此网站

**www.google.com** 意外终止了连接。

请试试以下办法：

- [检查网络连接](#)
- [检查代理服务器和防火墙](#)
- [运行 Windows 网络诊断](#)

ERR\_CONNECTION\_CLOSED

重新加载

详细信息

**Questions?**